SF 298 MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1996	

4. TITLE AND SUBTITLE

Principal Component and Factor Analyses

5. FUNDING NUMBERS

DAAH04-96-1-0082

6. AUTHOR(S)

C.R. Rao

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)

Center for Multivariate Analysis
417 Classroom Building
Penn State University
University Park, PA 16802

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

ARO 35578.2-MA-SDI

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

19960909 121

13. ABSTRACT *(Maximum 200 words)*

Principal component (PCA) and factor analyses (FA) are exploratory multivariate techniques used in studying the covariance (or correlation) structure of measurements made on individuals. The methods have been used by applied research workers in a variety of ways, from reducing high dimensional data to few functions of variables carrying the maximum possible information, grouping of similar measurements and detecting multicollinearity, to graphical representation of high dimensional data in lower dimensional spaces to visually examine to scatter of the data and detection of outliers.

The computations involved in these methods and the interpretation of results in different situations are discussed. The difference between PCA and FA, and the need to choose the appropriate technique in the analysis of given data are stressed.

It is shown that there is a close similarity between the growth curve models used in biometric studies and the arbitrage pricing theory model recently introduced in financial statistics.

14. SUBJECT TERMS	15. NUMBER IF PAGES
Arbitrage pricing theory model, Correspondence analysis, Cluster analysis, Factor analysis, Growth curve model, Prediction, Principal component analysis, Outlier detection.	20
	16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

PRINCIPAL COMPONENT AND FACTOR ANALYSES

C.R. Rao

Technical Report 96-04

July 1996

Center for Multivariate Analysis
417 Classroom Building
Penn State University
University Park, PA 16802

Research sponsored by the Army Research Office under Grant DAAHO4-96-1-0082. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

PRINCIPAL COMPONENT AND FACTOR ANALYSES

C.Radhakrishna Rao
Department of Statistics
The Pennsylvania State University
University Park, PA 16802

ABSTRACT

Principal component (PCA) and factor analyses (FA) are exploratory multivariate techniques used in studying the covariance (or correlation) structure of measurements made on individuals. The methods have been used by applied research workers in a variety of ways, from reducing high dimensional data to few functions of variables carrying the maximum possible information, grouping of similar measurements and detecting multicollinearity, to graphical representation of high dimensional data in lower dimensional spaces to visually examine to scatter of the data and detection of outliers.

The computations involved in these methods and the interpretation of results in different situations are discussed. The difference between PCA and FA, and the need to choose the appropriate technique in the analysis of given data are stressed.

It is shown that there is a close similarity between the growth curve models used in biometric studies and the arbitrage pricing theory model recently introduced in financial statistics.

Key Words: Arbitrage pricing theory model, Correspondence analysis, Cluster analysis, Factor analysis, Growth curve model, Prediction, Principal component analysis, Outlier detection.

1. Introduction

Principal component and factor analyses (PCA and FA) are exploratory multivariate techniques used in studying the covariance (or correlation) structure of measurements made on individuals. The object may vary from reduction of high dimensional data by finding a few latent variables which explain the variations of or the associations between the observable measurements, grouping of similar measurements and detecting multicollinearity, to graphical representation of high dimensional data in lower dimensional spaces to visually examine the scatter of the data, and detection of outliers. PCA was developed by Pearson (1901) and Hotelling (1933); a general theory with some extensions and applications are given in Rao (1964). FA originated with the work of Spearman (1904) and developed by Lawley (1940) under the assumption of multivariate normality. A general theory of FA, under the title Canonical Factor Analysis (CFA), without any distributional assumptions was given by Rao (1955). Now there are a number of excellent full length monographs devoted to the computational aspects and uses of PCA and CA in social and physical scientific research. Reference may be made to Bartholomew (1987), Basilevsky (1994), Cattel (1978), Jackson (1991), Jolliffe (1986) to mention a few authors.

A technique related to PCA, when the measurements are qualitative, is correspondence analysis (CA), developed by Benzecri (1973) based on a method of scaling qualitative categories suggested by Fisher (1936). A monograph by Greenacre (1984) gives the theory and applications of CA in the analysis of contingency tables. A recent paper by Rao (1995) contains an alternative to CA, which seems to have some advantages over the earlier approach, for the same purpose CA is used.

In this paper a general survey is given of PCA and FA with some recent theoretical results and practical applications.

2. Principal Components

2.1 *The general problem*

The problem of principal components can be stated in a very general set up as follows. Let $\mathbf{x}$ be a p-vector variable and y be a $\mathbf{q}$-vector variable, where some components of $\mathbf{x}$ and $\mathbf{y}$ may be the same. We want to replace $\mathbf{y}$ by $\mathbf{z} = A\mathbf{y}$ where A is an $r \times q$ matrix and $r < q$ in such a way that the loss in predicting $\mathbf{x}$ by using $\mathbf{z}$ instead $\mathbf{y}$ is as minimal as possible. If

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \tag{2.1}$$

is the covariance matrix of $\mathbf{x}$ and $\mathbf{y}$, then the covariance matrix of the errors in predicting $\mathbf{x}$ by $\mathbf{z} = A\mathbf{y}$ is

$$W = \Sigma_{11} - \Sigma_{12}A'(A\Sigma_{22}A')^{-1}A\Sigma_{21}. \tag{2.2}$$

We choose A such that $\|W\|$, for a suitably chosen norm, is small. If we choose $\|W\| = \operatorname{tr} W$, then the optimum choice is

$$A_* = \arg\max_A \operatorname{tr}\Sigma_{12}A'(A\Sigma_{22}A')^{-1}A\Sigma_{21}.$$

The maximum is attained at

$$A_*' = (C_1 : \ldots C_r) \tag{2.3}$$

where $C_1, \ldots, C_r$ are the r eigen vectors associated with the first r eigen values $\lambda_1^2 \geq \lambda_2^2 \geq \ldots \geq \lambda_r^2$ of $\Sigma_{21}\Sigma_{12}$ with respect to Σ_{22}, i.e., the eigen vectors and values are those arising out of the determinental equation

$$|\Sigma_{21}\Sigma_{12} - \lambda^2\Sigma_{22}| = 0. \tag{2.4}$$

The relative loss of information in using $\mathbf{z}_* = A_*\mathbf{y}$ for predicting $\mathbf{x}$ is

$$\begin{aligned} &\operatorname{tr}(\Sigma_{11} - (A_*\Sigma_{22}A_*')^{-1}A_*\Sigma_{21}\Sigma_{12}A_*')/\operatorname{tr}\Sigma_{11} \\ &\qquad = 1 - \frac{\lambda_1^2 + \ldots + \lambda_r^2}{\operatorname{tr}\Sigma_{11}}. \end{aligned} \tag{2.5}$$

We consider some special choices of $\mathbf{x}$ and $\mathbf{y}$ and derive the optimal transformation A as characterized in (2.3).

2.2 *The choice* $\mathbf{x} = \mathbf{y}$

The special choice, $\mathbf{x} = \mathbf{y}$, leads to the usual principal components $C_1'\mathbf{x}, \ldots, C_r'\mathbf{x}$, where $C_1, \ldots, C_r$ are the first eigen vectors associated with the first eigen values $\lambda_1^2 \geq \ldots \geq \lambda_r^2$ of the determinantal equation $|\Sigma_{11} - \lambda I| = 0$. In such a case, the loss of information (2.5) is

$$1 - \frac{\lambda_1^2 + \ldots + \lambda_r^2}{\lambda_1^2 + \ldots + \lambda_p^2} = \frac{\lambda_{r+1}^2 + \ldots + \lambda_p^2}{\lambda_1^2 + \ldots + \lambda_p^2} \tag{2.6}$$

usually expressed as a percentage. The choice of r is determined by the magnitude of (2.6).

In practice, we have to estimate λ_i^2 and C_i from a sample of n independent observations on the p-vector random variable $\mathbf{x}$, which we denote by the $p \times n$ matrix

$$X = (\mathbf{x}_1 : \ldots : \mathbf{x}_n). \tag{2.7}$$

An estimate of Σ_{11} is

$$S = (n-1)^{-1}X(I - \frac{1}{n}\mathbf{e}\mathbf{e}')X'$$

where $\mathbf{e}$ is an n-vector of unities. The estimates ℓ_i of λ_i and $\mathbf{c}_i$ of C_i are obtained from the spectral decomposition

$$S = \ell_1^2\mathbf{c}_1\mathbf{c}_1' + \ldots + \ell_p^2\mathbf{c}_p\mathbf{c}_p'. \tag{2.8}$$

The principal components of the observations on the i-th individual are then

$$\mathbf{q}_i = (\mathbf{c}_1'\mathbf{x}_i, \ldots, \mathbf{c}_p'\mathbf{x}_i)' \tag{2.9}$$

In the sequel, we denote

$$
\begin{aligned}
&s_{ii} = \text{the } i\text{-th diagonal element of } S, \\
&\mathbf{c}_j = (c_{j1}, \ldots, c_{jp})', \; j = 1, \ldots, p, && (2.10.1) \\
&\hat{c}_{ji} = \ell_j c_{ji}, \; i = 1, \ldots, p, && (2.10.2) \\
&\mathbf{q}_i = (q_{i1}, \ldots, q_{ip})', \; i = 1, \ldots, n, && (2.11.1) \\
&\hat{q}_{ij} = \ell_j^{-1} q_{ij}, \; i = 1, \ldots, n. && (2.11.2)
\end{aligned}
$$

It may be noted that the vectors $\mathbf{c}_i$ and $\mathbf{q}_i$ (apart from a translation of coordinates) can be obtained in one step from the singular value decomposition (SVD)

$$X(I - \frac{1}{n}\mathbf{e}\mathbf{e}') = \ell_1 \mathbf{c}_1 \mathbf{q}_1' + \ldots + \ell_p \mathbf{c}_p \mathbf{q}_p'. \qquad (2.12)$$

2.3 *Interpretation of principal components*

For an interpretation of principal components in terms of the influence of the original measurements on them, we need the following computations as exhibited in Table 1.

The magnitudes of the correlations in Table 1 indicate how well each variable is represented in each PC and overall in the first r PC's (judged by the values of R_i^2). The values of R_i^2 computed for $r = 1, 2, \ldots$ enable us to decide on r, the number of PC's to be chosen. If for some r, the values of R_i^2 are high except for one value of i, say j, then we may decide to include x_j along with $z_1, \ldots, z_r$ or add other PC's where x_j is well represented.

Table 1

original variable	correlation with principal component			multiple correlation of x_i on $z_1, \ldots, z_r$
	z_1	$\ldots$	z_p	
x_1	$\hat{c}_{11}/\sqrt{s_{11}}$	$\ldots$	$\hat{c}_{p1}/\sqrt{s_{11}}$	$\sum_{j=1}^{r} \hat{c}_{j1}^2/s_{11} = R_1^2$
.	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_p	$\hat{c}_{1p}/\sqrt{s_{pp}}$	$\ldots$	$\hat{c}_{pp}/\sqrt{s_{pp}}$	$\sum_{j=1}^{r} \hat{c}_{jp}^2/s_{11} = R_p^2$

2.4 *Graphical display of data*

To represent the individuals in terms of the original measurements we need a p-dimensional space. But for visual examination, we need a plot of the individuals in a two or a three dimensional space, which reflects the configuration of the individuals in the p-space (distances between individuals) to the extent possible. For this purpose, we use the PC's either as in (2.11.1) or in the standardized form [SPC as in (2.11.2)]. The full set of new coordinates in different dimensions from which first few may be selected is displayed in Table 2.

Table 2

individuals	dim 1		dim 2		...	dim p	
	PC	SPC	PC	SPC	...	PC	SPC
1	q_{11}	$\hat{q}_{11}$	q_{12}	$\hat{q}_{12}$	...	q_{1p}	$\hat{q}_{1p}$
2	q_{21}	$\hat{q}_{21}$	q_{22}	$\hat{q}_{22}$	...	q_{2p}	$\hat{q}_{2p}$
$\vdots$							
n	q_{n1}	$\hat{q}_{n1}$	q_{n2}	$\hat{q}_{n2}$	...	q_{np}	$\hat{q}_{np}$
Variance	ℓ_1^2	1	ℓ_2^2	1	...	ℓ_p^2	1

If we plot the individuals in the first $r(< p)$ dimensions using the coordinates $q_{i1}, \ldots, q_{ir}$ for the i-th individual, then the Euclidean distance between the individuals i and j in such a plot will be an approximation to the Euclidean distance in the full p-space

$$d_{ij} = [(\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j)]^{1/2}.$$

On the other hand, if we plot the individuals in the first $r(< p)$ dimensions using the coordinates $\hat{q}_{i1}, \ldots, \hat{q}_{ir}$, then the Euclidean distance between individuals i and j in such a plot will be an approximation to the Mahalanobis distance in the p-space

$$d_{ij} = [(\mathbf{x}_i - \mathbf{x}_j)'S^{-1}(\mathbf{x}_i - \mathbf{x}_j)]^{1/2}.$$

In practice, one may have to choose the appropriate distance we want to preserve in the reduced space. Usually, two or three dimensional plots may suffice to capture the original configuration. If more than three dimensions are necessary, other graphical displays for visualizing higher dimensional plots may be used. See for instance the paper by Wegman, Carr and Luo (1993).

We can also represent the variables in a lower dimensional space to provide a visual examination of the associations between them. The full set of coordinates for this purpose is given as in Table 3.

Table 3

variables	coordinates			
1	$\hat{c}_{11}$	$\hat{c}_{21}$	...	$\hat{c}_{p1}$
2	$\hat{c}_{12}$	$\hat{c}_{22}$	...	$\hat{c}_{p2}$
$\vdots$				
p	$\hat{c}_{1p}$	$\hat{c}_{2p}$	...	$\hat{c}_{pp}$

Let us denote the vector connecting the points representing the i-th individual in the r-dimensional space to the origin by $\mathbf{v}_i$. Then $\mathbf{v}_i'\mathbf{v}_i$ is a good approximation to s_{ii}, the variance of the i-th variable and the cosine of the angle between the vectors $\mathbf{v_i}$ and $\mathbf{v_j}$ will be a good approximation of the correlation between the i-th and j-th variables.

2.5 *Analysis of residuals and detection of outliers*

If we retain the first r PC's, we can compute the error in the approximation $\hat{\mathbf{x}}_i$ to $\mathbf{x}_i$, the p-vector of measurements on the i-th individual, by

$$\mathbf{x} - \hat{\mathbf{x}}_i = (\mathbf{c}_{r+1}\mathbf{c}_{r+1}' + \ldots + \mathbf{c}_p\mathbf{c}_p')\mathbf{x}$$

and an overall measure of difference is

$$d_i^2 = (\mathbf{x}_i - \hat{\mathbf{x}}_i)'(\mathbf{x}_i - \hat{\mathbf{x}}_i) = q_{ir+1}^2 + \ldots + q_{ip}^2.$$

If some d_i^2 is large compared to the others, we have an indication that $\mathbf{x}_i$ may be an outlier.

Note 1. The PC's are not invariant for linear transformations of the original variables. For instance, if the original variables are scaled by different numbers or if they are rotated by a linear transformation, the PC's will be different. This suggests that an initial decision has to be made on transforming the original measurements to a new set and then extracting the PC's. The recommendation usually made is to scale the measurements by the inverse of the standard deviations, which is equivalent to finding the PC's based on the correlation matrix rather than the covariance matrix.

Note 2. There are tests available on the eigen values and eigen vectors of a covariance matrix when the original measurements have a multivariate normal distribution. [See Chapter 4 of Basilevsky (1994)]. In practice, it may be necessary to test for normality of the original measurements if these tests are to be applied. It may be useful to try transformations of the measurements by using the Box-Cox family of transformations to induce normality if necessary. Several computer programs allow for this option. In such a case, we will be computing the PC's of transformed variables.

Note 3. In some problems such as the analysis of growth curves, the PC's are computed from the matrix $S = XX'$ without making correction for the mean. The references to such methods are Rao (1958, 1987).

Note 4. It has been suggested by Jolicoeur and Mosimann (1960) that the first principal component, which has the maximum variance, may be interpreted as a size factor provided all the coefficients are positive, and other principal components with positive and negative coefficients as shape factors. A justification for such an interpretation may be given as follows. Consider the i-th variable x_i in $\mathbf{x}$ and the j-th PC, $\mathbf{c_j'x}$ of $\mathbf{x}$. The regression of x_i on $\mathbf{c_j'x}$ is c_{ji} the i-th element in the j-th eigen vector $\mathbf{c_j}$. Now a unit increase in $\mathbf{c_j'x}$ produces on the average an increase c_{ji} in x_i. If all the elements in $\mathbf{c_j}$ are positive, a unit increase in $\mathbf{c_j'x}$ increases the value of each of the measurements, in which case $\mathbf{c_j'x}$ may be described as a size factor. If some coefficients are positive and others are negative, then an increase in $\mathbf{c_j'x}$ increases the values of some measurements and decreases the values of the others, in which case $\mathbf{c_j'x}$ may be interpreted as a shape factor.

It may be of interest to note that if all the original measurements are non-negative, then the first PC of the uncorrected sum of squares and products matrix will have all its coefficients non-negative.

Note 5. Another particular case of the general problem stated in Section 2.1 is when $\mathbf{x}$ and $\mathbf{y}$ are completely different sets of variables. Such a situation arises when we have a large number of what are called instrumental variables represented by $\mathbf{y}$, and we wish to predict each dependent variable in the set $\mathbf{x}$ using certain linear functions $\mathbf{y}$. Such a procedure may be more economical and sometimes more efficient due to multicollinearity in $\mathbf{y}$.

2.6 *Principal components of* $\mathbf{x}$ *uncorrelated with concomitant variables* $\mathbf{z}$

In some problems it is of interest to find the principal components of a p-vector $\mathbf{x}$ uncorrelated with a q-vector of concomitant variables $\mathbf{z}$. Let

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \tag{2.13}$$

denote the covariance matrix of $(\mathbf{x}', \mathbf{z}')'$ in the partitioned form. We need k principal components $\mathbf{L_1'x}, \ldots, \mathbf{L_k'x}$ such that $\mathbf{L_i'L_i} = 1$, $\mathbf{L_i'L_j} = 0$ and $\text{cov}(\mathbf{L_i'x}, \mathbf{z}) = \mathbf{L_i'}\Sigma_{12} = 0$, $i, j =$

$1, \ldots, k$ and

$$\mathbf{L}_1'\Sigma\mathbf{L}_1 + \ldots + \mathbf{L}_k'\Sigma\mathbf{L}_k \tag{2.14}$$

is a maximum. It is shown in Rao (1964), that the optimum choice of $L_1, \ldots, L_k$ are the first k right eigen vectors of the matrix

$$(I - \Sigma_{12}(\Sigma_{21}\Sigma_{12})^{-1}\Sigma_{21})\Sigma_{11}. \tag{2.15}$$

As an application, let us consider a p-vector time series representing some blocks of economic transactions considered by Stone (1947).

Economic transactions	Time periods 1	2	...	T
1	x_{11}	x_{12}	...	x_{1T}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
p	x_{p1}	x_{p2}	...	x_{pT}
Concomitants functions of time				
linear	1	2	...	T
quadratic	1	2^2	...	T^2

We compute the $(p+2)$ order covariance matrix arising out of the main variables and concomitants, considering T as the sample size,

$$\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \tag{2.16}$$

where S_{11} is of order $p \times p$, S_{12} of order $p \times 2$ and S_{22} of order 2×2.

The necessary number of right eigen vectors of

$$(I - S_{12}(S_{21}S_{12})^{-1}S_{21})S_{11} \tag{2.17}$$

provide principle components of x unaffected by linear and quadratic trends of the transactions over time. Elimination of lower order or higher order trends is possible by suitably choosing the concomitant variables as powers of time.

Stone (1947) considered the above problem of isolating linear functions of $\mathbf{x}$ which have an intrinsic economic significance from those which represent trend with time and those which measure random errors. For this purpose he computed the covariance matrix of $\mathbf{x}$ variables alone and found the PC's using the eigen vectors of the S_{11} part of the matrix without any reference to the time factor. The problem was then posed as that of identifying the dominant PC which accounted for a large variance. This was interpreted as linear trend and

other PC's were interpreted in economic terms. It is believed that the method suggested of obtaining the PC's using the matrix (2.17) is more flexible and provides a better technique of eliminating trend of any order and providing linear functions with intrinsic economic significance.

3. Model Based Principal Components

3.1 *An analogy with the factor analytic model*

Let us suppose that the measurement p-vector $\mathbf{x}_i$ on individual i can be expressed as

$$\mathbf{x}_i = \alpha + A\mathbf{f_i} + \epsilon_i,\ i = 1, \ldots, n \tag{3.1}$$

where α is a p-vector and A is $p \times r$ matrix common to all individuals, $\mathbf{f_i}$ is an r-vector specific to individual i, and ϵ_i is a random variable such that $E(\epsilon_i) = 0$, and $V(\epsilon_i) = \sigma^2$ for $i = 1, \ldots, n$. The model (3.1) is analogous to the FA model except that in FA the covariance matrix of ϵ_i is diagonal with possibly different elements (see Section 4 of the paper). The problem we consider is one of estimating $A, \mathbf{f_i}, \ldots, \mathbf{f_n}$ and σ^2 from the model (3.1). Note that the solution is not unique unless we impose certain restrictions such as that the columns of A are orthonormal. We can write the joint model (3.1) as

$$X = \alpha\mathbf{e}' + AF + E \tag{3.2}$$

where $X = (\mathbf{x_1} : \ldots : \mathbf{x_n})$ is $p \times n$ matrix, $\mathbf{e}$ is an n-vector of unites, and F is $r \times n$ matrix. We may estimate α, A and F by minimizing

$$\|X - \alpha\mathbf{e}' - AF\| \tag{3.3}$$

for an appropriately chosen norm. The choice of Frobenius norm leads to an extended method of least squares where the expression

$$\sum_{i=1}^{n}(\mathbf{x}_i - \alpha - A\mathbf{f_i})'(\mathbf{x}_i - \alpha - A\mathbf{f_i}) \tag{3.4}$$

is minimized with respect to α, A and $\mathbf{f_i}, \ldots, \mathbf{f_n}$. One possible solution (see Rao (1995)) is

$$\hat{\alpha} = \bar{\mathbf{x}},\ \hat{A} = (\mathbf{c_1} : \ldots : \mathbf{c_r})$$
$$\hat{\mathbf{f}}_\mathbf{i} = \hat{A}'(\mathbf{x}_i - \bar{\mathbf{x}}) \tag{3.5}$$

where $\mathbf{c}_1, \ldots, \mathbf{c}_r$ are the first r eigen vectors of $S = X(I - \frac{1}{n}\mathbf{e}\mathbf{e}')X'$. Then $\hat{\mathbf{f}}_i$ is the vector of r PC's for the individual i. We thus have the same solution as that discussed in Sections 2.2 - 2.5. An estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{n-1}{(n-r-1)(p-r)}(\ell_{r+1}^2 + \ldots + \ell_p^2) \tag{3.6}$$

where $\ell_{r+1}^2, \ldots, \ell_p^2$ are the last $(p-r)$ eigen values of S.

In some problems, it may be appropriate to consider $\mathbf{f}_i$ in the model (3.1) as a random variable with the identity I as covariance matrix. In such a case

$$E(S) = AA' + \sigma^2 I, \tag{3.7}$$

an estimate of A is

$$\hat{A} = (\ell_1 \mathbf{c}_1 : \ldots : \ell_r \mathbf{c}_r), \tag{3.8}$$

and an estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{n-1}{(n-r-1)(p-r)}(\ell_{r+1}^2 + \ldots + \ell_p^2) \tag{3.9}$$

which are the same as in (3.6) except for scaling factors. If it is desired to estimate (predict) $\mathbf{f}_i$, one may use the regression of $\mathbf{f}_i$ on $\mathbf{x}_i$ which is of the form

$$\hat{\mathbf{f}}_i = \hat{A}'(\hat{A}\hat{A}' + \hat{\sigma}^2 I)^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}) \tag{3.10}$$

and differs from the expression (3.5). A similar situation arises when we want to estimate the parameters simultaneously from several linear models having the same design matrix. Reference may be made to Rao (1975) for a discussion of such a problem.

3.2 *Regression problem based on a PC model*

We have n independent observations on a $(p+1)$-vector random variable $(y, \mathbf{x})$, where $\mathbf{x}$ is a p-vector and y is a scalar,

$$(y_1, \mathbf{x}_1), \ldots, (y_n, \mathbf{x}_n) \tag{3.11}$$

and only $\mathbf{x}_{n+1}$ for the $(n+1)$-th sample. The problem is to predict y_{n+1} the unobserved value, under the PC model

$$\mathbf{x}_i = \alpha_1 + A\mathbf{f}_i + \epsilon_i \tag{3.12}$$

$$y_i = \alpha_2 + \mathbf{b}'\mathbf{f}_i + \eta_i \tag{3.13}$$

$$i = 1, \ldots, n+1$$

where $\text{cov}(\epsilon_i, \eta_i) = 0$, $\text{cov}(\epsilon_i) = \sigma^2 I$, $V(\eta_i) = \sigma_0^2$, and the rest of the assumptions are the same as in the model (3.1). The above problem was considered in a series of papers (see Rao (1975, 1976, 1978, 1987) and Rao and Boudreau (1985)). Recently, the model (3.12 - 3.13) is used in the development of partial least squares (see Helland (1988) and the references there in).

There are several possible approaches to the problem.

1) Let $\hat{\mathbf{f}}_1, \ldots, \hat{\mathbf{f}}_{n+1}$ be the estimates of $\mathbf{f}_1, \ldots, \mathbf{f}_{n+1}$ using the observational equations (3.12) only. Then find estimates $\hat{\alpha}_2$ and $\hat{\mathbf{b}}$ of α_2 and $\mathbf{b}$, using the first n observational equations of (3.13) and assuming $\hat{\mathbf{f}}_1, \ldots, \hat{\mathbf{f}}_n$ as known, by the usual least squares method. Finally predict y_{n+1} by the formula

$$\hat{y}_{n+1} = \hat{\alpha}_2 + \hat{\mathbf{b}}'\hat{\mathbf{f}}_{n+1}. \tag{3.14}$$

2) Let $\hat{\alpha}_1, \hat{\alpha}_2, \hat{A}$ and $\hat{\mathbf{b}}$ be the estimates of α_1, α_2, A and $\mathbf{b}$ using the first n observational equations in (3.12) and (3.13). Then estimate $\mathbf{f}_{n+1}$ using the equations

$$\mathbf{x}_{n+1} = \hat{\alpha}_1 + \hat{A}\mathbf{f}_{n+1} + \epsilon_{n+1} \tag{3.15}$$

assuming $\hat{\alpha}_1$ and $\hat{A}$ as known, by the least squares method. If $\hat{\mathbf{f}}_{n+1}$ is the estimate of $\mathbf{f}_{n+1}$, then y_{n+1} is predicted by

$$\hat{y}_{n+1} = \hat{\alpha}_2 + \hat{\mathbf{b}}'\hat{\mathbf{f}}_{n+1}. \tag{3.16}$$

3) Substitute a value say y for y_{n+1} to make the equations (3.12 - 3.13) complete. Then find the singular value decomposition of the partitioned matrix

$$\begin{pmatrix} \mathbf{x}_1 : \ldots : \mathbf{x}_n & \mathbf{x}_{n+1} \\ y_1 : \ldots : y_n & y \end{pmatrix} (I - (n+1)^{-1}\mathbf{e}\mathbf{e}') = \ell_1 \mathbf{c}_1 \mathbf{q}_1' + \ldots + \ell_{p+1}\mathbf{c}_{p+1}\mathbf{q}_{p+1}'$$

where ℓ_i depend on y, and compute

$$S_r(y) = \ell_{r+1}^2(y) + \ldots + \ell_{p+1}^2(y). \tag{3.17}$$

Finally predict y_{n+1} as the value of y which minimizes (3.17). The solution may be obtained graphically or by an iterative algorithm as described in Rao and Boudreau (1985).

4) Another method is to consider $\mathbf{f}_i$ as a random variable with zero mean vector and covariance matrix Γ. Then

$$\text{cov} = \begin{pmatrix} \mathbf{x}_i \\ y_i \end{pmatrix} = \begin{pmatrix} A\Gamma A' + \sigma_1^2 I & A\Gamma b \\ b'\Gamma A' & b'\Gamma b + \sigma_2^2 \end{pmatrix}. \tag{3.18}$$

Using (3.12) and the first n observational equations in (3.13), obtain the estimates of $A, \Gamma, \mathbf{b}, \sigma_1^2$ and σ_2^2. Methods described by Bentler (1983), Sörbom (1974) and Rao (1983, 1985) may be used for this purpose. Then y_{n+1} may be predicted by

$$\hat{y}_{n+1} = \bar{y} + \mathbf{b}'\Gamma A'(A\Gamma A' + \sigma_1^2 I)^{-1}(\mathbf{x_{n+1}} - \bar{\mathbf{x}}) \tag{3.19}$$

where $\bar{y} = n^{-1}\Sigma y_i$, $\bar{\mathbf{x}} = (n+1)^{-1}\Sigma \mathbf{x}_i$ and for $\mathbf{b}, \Gamma, A$ and σ_1^2 their estimates are substituted.

4. Factor Analysis

4.1 *General discussion*

In FA, a p vector variable $\mathbf{x}$ is endowed with a stochastic structure

$$\mathbf{x} = \alpha + A\mathbf{f} + \epsilon \tag{4.1}$$

where α is a p-vector and A is $p \times r$ matrix of parameters, $\mathbf{f}$ is an r-vector of latent variables called common factors and ϵ is a p-vector of variables called specific factors, with the following assumptions:

$$\begin{aligned} E(\epsilon) = 0,\ \mathrm{cov}(\epsilon) = \Delta \quad &\text{a diagonal matrix} \\ E(\mathbf{f}) = 0,\ \mathrm{cov}(\mathbf{f}, \epsilon) = 0,\ \mathrm{cov}(\mathbf{f}) &= I. \end{aligned} \tag{4.2}$$

As a consequence of (4.2), we have

$$\Sigma = \mathrm{cov}(\mathbf{x}) = AA' + \Delta. \tag{4.3}$$

Note that (4.3) reduces to the PC model considered in (3.1) when $\Delta = \sigma^2 I$. The problems generally discussed in FA, on the basis of n independent observations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ made on $\mathbf{x}$, are:

1) What is the minimum r for which the representation (4.3) holds?
2) How do we estimate A, called the matrix of factor loadings?
3) How do we interpret the factors?
4) How do we estimate $\mathbf{f}$ for a given individual given the observable $\mathbf{x}$?

It may be noted that the equation (4.3) does not ensure the existence of a unique A even for a given r and so also $\mathbf{f}$ in (4.1). However, the object is to obtain any particular solution, and consider transformations of A and $\mathbf{f}$ for an interpretation. References to a discussion of non-identifiability of A and $\mathbf{f}$ and rotation of factors are Basilevsky (1994, pp.355-360, 402-404), Jackson (1991, pp.393-396), Jolliffe (1986, pp.117-118).

Denoting $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, we compute

$$\bar{\mathbf{x}} = n^{-1} X \mathbf{e}$$
$$S = (n-1)^{-1} X (I - n^{-1} \mathbf{e}\mathbf{e}') X'$$

as estimates of α and Σ. Then estimate A and Δ starting with S. The most commonly used method is maximum likelihood (ML) under the assumption of multivariate normality of the vector variable $\mathbf{x}$. There are a number of computer packages for the estimation of r, the number of factors, A, the matrix of factor loadings and Δ, the matrix of specific factor variances. (See for instance SPSS, SAS, OSIRIS, BMD, COFAMM etc., which also offer alternatives other than ML estimates and also compute rotations of factor loadings for interpretation). Let us denote the ML estimates of A and Δ by $\hat{A}$ and $\hat{\Delta}$.

The likelihood ratio test criterion for testing the hypothesis that there r common factors is

$$-(n-1)\log \frac{|S|}{|\hat{A}\hat{A}' + \hat{\Delta}|} \tag{4.5}$$

which is asymptotically distributed as χ^2 on $[(p-r)^2 - p - r]/2$ degrees of freedom in large samples. This is valid under the assumption of multivariate normality. A slight improvement to the χ^2 approximation is obtained by replacing the multiplier $(n-1)$ in (4.5) by

$$n - 1 - \frac{2p+5}{6} - \frac{2r}{3}. \tag{4.6}$$

An alternative method called canonical factor analysis (CFA) for the estimation A and Δ is developed by Rao (1955) without making any distributional assumptions. The solution turns out to be same as the ML estimate. However, the χ^2-test of (4.5) requires the assumption of multivariate normality.

A general recommendation is to test for multivariate normality based on the observed data $\mathbf{x}_1, \ldots, \mathbf{x}_n$ using some of the techniques available in computer packages. Some references to a discussion of tests of normality are Basilevsky (1994, Section 4.6.2) and Gnanadesikan (1977, Section 5.4.2). It may also be worthwhile making transformations of variables to achieve normality. But in such a case the factor structure has to be imposed on transformed variables.

It may be noted that unlike PCA, the FA is invariant under scaling of variables, if one uses scale free extraction methods such as the ML and CFA. In these cases, one can use the covariance or the correlation matrix to start with. If the covariance matrix is used and scales vary very widely, scale factors will complicate interpretation of results. In such a case, there is some advantage in using the correlation matrix. The covariance matrix is preferable when comparison of factor structures between groups is involved (see Sörbom (1974)).

4.2 *Estimation of factor scores*

Using the estimates $\hat{A}$ and $\hat{\Delta}$ of A and Δ in the representation of Σ, we can estimate the factor score $\mathbf{f}_i$ of the i-th individual with measurements $\mathbf{x}_i$ by

$$\hat{\mathbf{f}}_i = \hat{A}'(\hat{A}\hat{A}' + \hat{\Delta})^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}),\ i = 1, \ldots, n. \tag{4.7}$$

The expression (4.7) is simply the regression of $\mathbf{f}_i$ on $\mathbf{x}_i$ with the estimates substituted for the unknowns. There are other expressions suggested for the estimates of factor scores (see Jackson (1991, p.409)).

4.3 *Prediction problem*

We consider a $p+1$ variable $(\mathbf{x}, y)$ with the factor structure

$$\begin{aligned} \mathbf{x} &= \alpha + A\mathbf{f} + \epsilon \\ y &= \beta + \mathbf{a}'\mathbf{f} + \eta \end{aligned} \tag{4.8}$$

where β is a scalar, $\mathbf{a}$ is an r-vector and η is such that $E(\eta) = 0$, $\mathrm{cov}(\epsilon, \eta) = 0$, $V(\eta) = \delta^2_{p+1}$. Suppose that we have observations $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ on n individuals and only $\mathbf{x}_{n+1}$ on an $(n+1)$-th individual. The problem is to predict y_{n+1}, given all the other observations. By considering the factor structure of the $(p+1)$-vector variable

$$\begin{pmatrix} \mathbf{x} \\ y \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} A \\ \mathbf{a}' \end{pmatrix} \mathbf{f} + \begin{pmatrix} \epsilon \\ \eta \end{pmatrix} \tag{4.9}$$

and using the observations $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ we estimate all the unknown parameters. Let $\hat{\alpha}, \hat{\beta}, \hat{A}, \hat{\mathbf{a}}, \hat{\Delta}$ and $\hat{\delta}^2_{p+1}$ be estimates of the corresponding parameters using the CFA or ML-method. Then the regression estimate of y_{n+1} given $\mathbf{x}_{n+1}$ is

$$\hat{y} = \hat{\beta} + \hat{\mathbf{a}}'\hat{A}'(\hat{A}\hat{A}' + \hat{\Delta})^{-1}(\mathbf{x}_{n+1} - \hat{\alpha}). \tag{4.10}$$

In this case, we are not utilizing the information provided by $\mathbf{x}_{n+1}$, on the parameters $\mathbf{a}$, A and Δ.

4.4 *What is the difference between PCA and FA?*

In PCA, we do not impose any structure on the p-vector random variable $\mathbf{x}$. Suppose that $E(\mathbf{x}) = 0$ and $\mathrm{cov}(\mathbf{x}) = \Sigma$. We wish to replace $\mathbf{x}$ by a smaller number of linear combinations

$\mathbf{y} = L'\mathbf{x}$ where L is $p \times r$ matrix of rank r. Then the predicted value of $\mathbf{x}$ given $\mathbf{y}$ (i.e., the regression of $\mathbf{x}$ on $\mathbf{y}$) is

$$\hat{\mathbf{x}} = \Sigma L(L'\Sigma L)^{-1}\mathbf{y} \tag{4.11}$$

and the covariance matrix of the residual $\mathbf{x} - \hat{\mathbf{x}}$ is

$$\Sigma - \Sigma L(L'\Sigma L)^{-1}L'\Sigma. \tag{4.12}$$

We wish to choose L to minimize a suitable norm of (4.12). The choice of Frobenius norm leads to the solution

$$L = (\mathbf{c}_1 : \ldots : \mathbf{c}_r) \tag{4.13}$$

where $\mathbf{c}_1, \ldots, \mathbf{c}_r$ are the first r eigen vectors of Σ in which case $L'\mathbf{x}$ represents the first r principal components as explained in Section 3. The aim is to account for the entire covariance matrix of $\mathbf{x}$, to the extent possible, in terms of a reduced number of variables.

In FA, we are fitting an expression of the type $AA' + \Delta$ to R, the correlation matrix of the p-vector variable $\mathbf{x}$. Since Δ is a diagonal matrix of free parameters, the matrix A is virtually determined by minimizing the differences between the off diagonal elements of AA' and R. Thus, the matrix of factor loadings is designed to explain the correlations between the observed variables. The variances in the variables unexplained by the factors, irrespective of their magnitudes, is characterized as specific variances. In PCA, the emphasis is more on explaining the overall variances arising out of common and specific factors. Thus, the objectives of PCA and CA are different and so are the solutions.

Note 1. Fitting an expression of the type $AA' + \Delta$ to R imposes an automatic upper bound to r, the number of factors. So, in a given situation, one is forced to interpret the data in terms of far fewer factors than those that may have influenced the data. In the CFA developed by the author (Rao (1955)), no limit is placed on the number of common factors, but the method allows for the requisite number of dominant factors to be extracted from the data. No fixed number of factors is postulated to begin with, and the problem is treated as one of estimation rather than testing of hypothesis on the number of factors.

Note 2. It may be of interest to note that in the formulation of the FA model, only the second order properties of the common and specific factors are used. However, if we demand independence of the distribution of all these variables, the problem becomes more complex as the following theorem proved in Rao (1969, 1973) shows.

Theorem: Let $\mathbf{x}$ be a p-vector random variable with a linear structure $\mathbf{x} = A\mathbf{y}$, where $\mathbf{y}$ is a q-vector of independent r.v.'s. Then $\mathbf{x}$ admits the decomposition

$$\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are independent, $\mathbf{x}_1$ has essentially a unique structure ($\mathbf{x}_1 = A_1\mathbf{y}_1$ with a unique A_1 apart from scaling and $\mathbf{y}_1$ as a vector of a fixed number of independent non-normal

variables) and $\mathbf{x}_2$ has a p-variate normal distribution with a non-unique linear structure ($\mathbf{x}_2 = B_2\mathbf{y}_2$ with B_2 not necessarily unique and $\mathbf{y}_2$ as a vector of independent univariate normal variables).

In view of this theorem, if some of the factors have a non-normal distribution, the uniqueness of A_1 automatically specifies a lower bound to the number of factor variables which may have no relationship with p. The limitations placed on the FA model by considering only second order properties of the variables involved need some investigation.

4.5 *The arbitrage pricing theory model (APT)*

The classical FA model is extended to a statistical model of the APT by Ross (1976), which is similar to the growth curve model of Rao (1958, equation 9, Section 3). Consider the usual FA model, using the notation used in the finance literature

$$\mathbf{R} = \mu + B\mathbf{f} + \mathbf{u} \tag{4.18}$$

where R denotes the N-vector of returns on N assets, $\mu = E(\mathbf{R})$, $E(\mathbf{f}) = 0$, $E(\mathbf{u}) = 0$, $E(\mathbf{fu}') = 0$, $\text{cov}(\mathbf{f}) = \Phi$ and $\text{cov}(\mathbf{u}) = \triangle$, a diagonal matrix. The matrix B of order $N \times k$ is the matrix of factor loadings. [In the earlier sections p is used for N and r for k]. From the assumptions made

$$\Sigma = \text{cov}(R) = B\Phi B' + \triangle \tag{4.19}$$

Now, we model μ as

$$\mu = R_f\mathbf{e} + B\lambda \tag{4.20}$$

where R_f is described as the riskless return on a riskless asset. The sample we have over T time periods is

$$(\mathbf{R}_1, R_{f1}), \ldots, (\mathbf{R}_\mathbf{T}, R_{fT}) \tag{4.21}$$

where in (4.21), R_f is known and varies over time and λ is k-vector of unknown parameters called the factor premiums. Writing $\mathbf{r_t} = \mathbf{R_T} - R_{fT}\mathbf{e}$, we can write the model for the t-th observation as

$$\mathbf{r_t} = B(\mathbf{f_t} + \lambda) + \mathbf{u_t},\ t = 1, \ldots, T \tag{4.22}$$

which is exactly the model considered in Rao (1958). The marginal model for $\mathbf{r_t}$ is

$$\mathbf{r_t} = B\lambda + \mathbf{v_t},\ t = 1, \ldots, T \tag{4.23}$$

with $\text{cov}(\mathbf{v_t}) = \Sigma$. If B and Σ are known, the least squares estimate of λ is

$$\hat{\lambda} = (B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}\bar{\mathbf{r}} \tag{4.24}$$

where $\bar{\mathbf{r}} = T^{-1}(\mathbf{r}_1 + \ldots + \mathbf{r}_T)$. If B and Σ are not known, it is suggested by Roll and Ross (1980) and also Rao (1958) that they can be estimated by ML or an appropriate

nonparametric method considering the model (4.18) with unrestricted μ as discussed in section 4.2 of this article and substituted in (4.23). If multivariate normality is assumed for the distribution of $\mathbf{f}$ and $\mathbf{u}$ in the model (4.18), it is possible to write down the likelihood for all the unknown parameters B, λ, Φ and $\triangle$ based on the observations $\mathbf{r}_1, \ldots, \mathbf{r}_T$ and obtain the ML estimates for all the unknown parameters. We can then also apply likelihood ratio tests for the specification of Σ, i.e., for the number of factors, and the structure (4.20) on μ. Such a procedure is fully worked out in Christensen (...), where the method is applied to New York Stock Exchange data.

5. Conclusions

Both PCA and FA may be considered as multivariate methods for exploratory data analysis. The aim of both the analyses is to understand the structure of the data, through reducing the number of variables, which in some sense can replace the original data and which are easier to study through graphical representation and multivariate inference techniques. Some caution is necessary as there are many decisions to be made on the number of reduced variables and the criterion by which adequacy of the reduced set of variables in representing the whole set of original variables is judged.

Some practioners consider PCA and FA as alternative techniques of multivariate data analysis intended to answer the same questions. It is also claimed that each technique has evolved into a useful data - analytic tool and has become an invaluable aid to other statistical models such as cluster and discriminant analysis, least squares regression, graphical data displays, and so forth. As discussed in the present article, the purposes of reduction of data in PCA and FA are different. In PCA, the reduced data is intended to approximate, to the maximum possible extent, the dispersion of the original data in terms of the entire covariance matrix, while in FA, the emphasis is on explaining the correlations or association between the original variables. The objectives are different and a decision has to be made as to the appropriateness of PCA or FA in a particular situation and the purpose of data analysis. While the roles of PCA and FA in exploratory data analysis are clear, the exact uses of the estimated PC's and factors in inferential data analysis, or in planning further investigations do not seem to be satisfactorily laid out.

Some conditions under which the factor scores and principal components are close to each other have been given by Schneeweiss and Mathes (1955). It would be of interest to pursue such theoretical investigations and also examine in individual data sets the actual differences between principal components and factor scores.

6. References

1. Bartholomew, D.J. (1987). *Latent Variable Models and Factor Analysis*, Oxford University Press, New York.

2. Basilevsky, A. (1994). *Statistical Factor Analysis and Related Methods*, Wiley, New York.

3. Bentler, P.M. (1983). Some contributions to efficient statistics in structural models: specification and estimation of moment structures, *Psychometrika* 48, 493-517.

4. Benzecri, J.P. (1973). L'analyze des Donnéses, Tome II, *L'Analyse des Correspondences*, Dunod, Paris.

5. Cattel, R.B. (1978). *The Scientific Use of Factor Analysis in Behavioural and Life Science*, Plenum Press.

6. Christensen, B.J. (...). The likelihood ratio test of the APT with unobservable factors against the unrestricted factor model. ?????

7. Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems, *Ann. Eugen.* (London) 7, 179-188.

8. Gnanadesikan, R. (1977). *Methods for Statistical Analysis of Multivariate Observations*, Wiley, New York.

9. Greenacre, M.J. (1984). *Theory and Applications of Correspondence Analysis*, Academic, London.

10. Helland, I.S. (1988). On the structure of partial least squares regression, *Commun. Statist. Simula.* 17, 581-607.

11. Hotelling, H. (1933). Analysis of a complex of statistical variable into principal components, *Psychometrika*, 1, 27-35.

12. Jackson, J.E. (1991). *A User's Guide to Principal Components*, Wiley, New York.

13. Jolicoeur, P. and Mosiman, J.E. (1960). Size and shape variation in the painted turtle, a principal component analysis. *Growth* 24, 339-354.

14. Joliffe, I.T. (1986). *Principal Component Analysis*, Springer-Verlag, New York.

15. Lawley, D.N. (1940). The estimation of factor loadings ny the method of maximum likelihood, *Proc.Roy.Soc.Edinburgh* (A), 60, 64-82.

16. Pearson, K. (1901). On lines and planes of closest fit to a system of points in space, *Philosophical Magazine* 2, 6-th Series, 557-572.

17. Rao, C.R. (1955). Estimation and tests of significance in factor analysis, *Psychometrika* 20, 93-111.

18. Rao, C.R. (1958). Some statistical methods for comparison of growth curves, *Biometrics* 14, 1-17.

19. Rao, C.R. (1964). The use and interpretation of principal component analysis in applied research, *Sankhyā* A 26, 329-358.

20. Rao, C.R. (1969). A decomposition theorem for vector variables with a linear structure, *Ann. Math. Statist.* 40, 1845-1849.

21. Rao, C.R. (1973). *Linear Statistical Inference and its Applications* (Second edition), Wiley, New York.

22. Rao, C.R. (1975). Simultaneous estimation of parameters in different linear models and applications to biometric problems, *Biometrics* 31, 545-554.

23. Rao, C.R. (1976). Prediction of future observations with special reference to linear models, in *Multivariate Analysis IV* (Ed. P.R. Krishnaiah), North Holland, 193-208.

24. Rao, C.R. (1983). Likelihood ratio tests for relationships between covariance matrices, in *Studies in Economics, Time Series and Multivariate Statistics* (Eds. S. Karlin, T. Amemiya and L.A. Goodman), Academic, New York, 529-543.

25. Rao, C.R. and Boudreau, R. (1985). Prediction of future observations in factor analytic type growth model, in *Multivariate Analysis VI* (Ed. P.R. Krishnaiah), Elsevier, Amsterdam, 449-466.

26. Rao, C.R. (1987). Prediction of future observations in growth curve models, *J. Statistical Science* 2, 434-471.

27. Rao, C.R. (1995). A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance, *Qüestiió* 19, 23-63.

28. Roll, R. and Ross, S.A. (1980). An empirical investigation of the arbitrage pricing theory, *J. of Finance* 35, 1073-1103.

29. Ross, S.A. (1976). The arbitrage theory of capital asset pricing, *J. of Economic Theory* 13, 341-360.

30. Schneeweiss, H. and Mathes, H. (1995). Factor analysis and principal components, *J. Multivariate Analysis* 55, 105-124.

31. Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British J. Math. Statist. Psych.* 27, 229-239.

32. Spearman, C. (1904). General intelligence, objectively determined and measured, *Am. J. Psych.* 15, 201-293.

33. Stone, R. (1947). An interdependence of blocks of transactions, *J. Roy. Statist. Soc.* (Supple), 8, 1-32.

34. Wegman, E.J., Carr, D.B., Luo, Q. (1993). Visualizing multivariate data, in *Multivariate Analysis: Future Directions* (Ed. C.R. Rao) North Holland, 423-466.